

Dancing with data



The data

Dataset with songs from 1950 until 2019

28.372 songs

27.614. songs in English

From many different artists and genres and including the lyrics

Source: Mendeley Data

Final dataset variables

- O Artist
- O Song name
- O Release Date
- O Lyrics
- O Genre
- O Topic



Dataset available at https://data.mendeley.com/datasets/3tgvbwxgr5/1

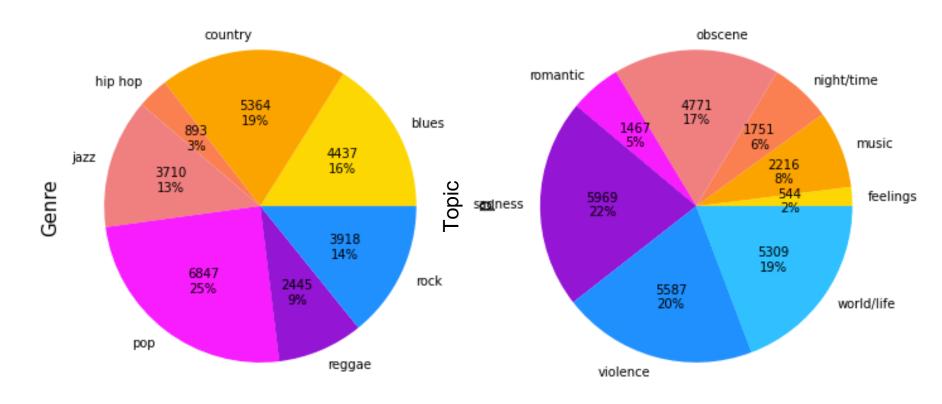


NLP - Text preprocessing



Genre & topic distribution



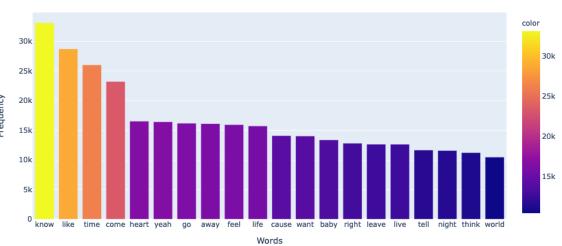




Most frequent words



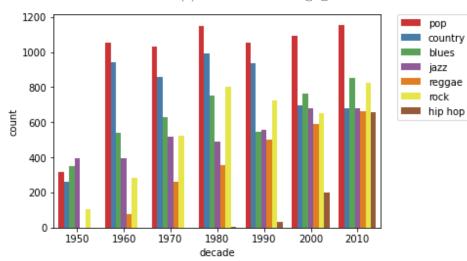
Top 20 word frequencies in the dataset

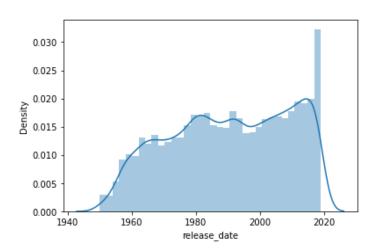




Genre evolution per decade

- Songs from 1950 until 2019
- Pop is the most popular genre in almost all decades
- Few songs in the 1950s less representative

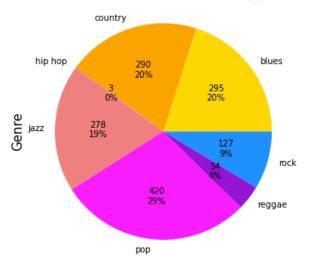


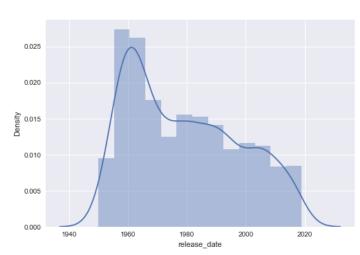




Romantic songs

- 1467 songs labelled as romantic
- Pop is the genre that most romantic songs contains
- Fewer romantic songs with time



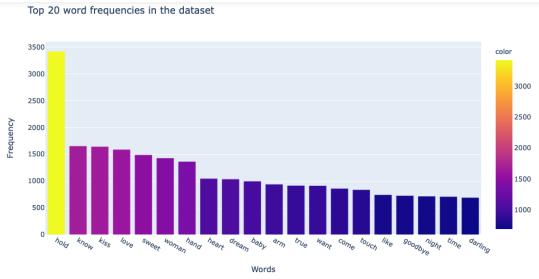




Romantic songs

• Specific vocabulary compared to the most frequent words in the whole dataset

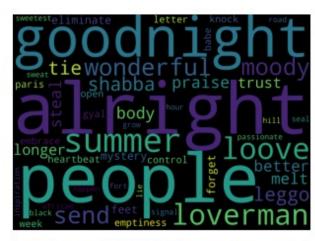




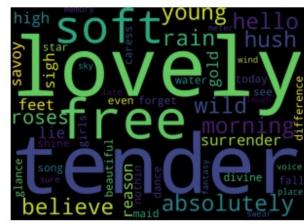


Topic modeling

Whole dataset (1950-2019)











Topic modeling – a closer look

1950-1975







1975-2000









2000-2019











Topic modeling – a closer look

1950-1975











1975-2000









2000-2019

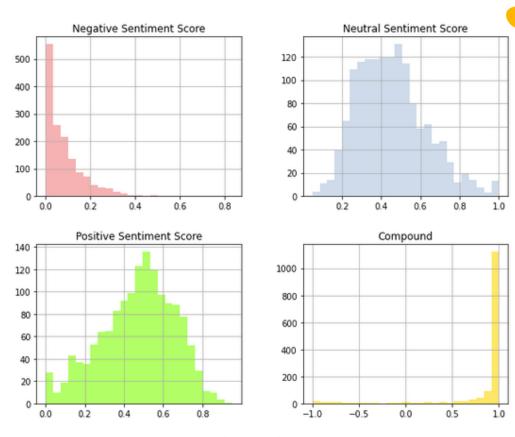






Sentiment Analysis of romantic lyrics

Sentiment analysis analysis











*

Song recommender

Based on text similarity (lyrics), using BoW and cosine similarity

1 get_recommendations('accept yourself')

	track_name	artist_name
16733	king of love	dave edmunds
13237	the sounds	ten years after
3715	someday	mariah carey
20792	whereyougonnago?	jitwam
10460	someday you'll call my name	hank williams
6141	we come running	youngblood hawke
10446	someday, somewhere, somehow	billy ray cyrus
18679	can it be done	weather report
2305	i love the sound of breaking glass	nick lowe
16520	toe cutter	thee oh sees

get_recommendations('radio ga ga')

	track_name	artist_name
26809	handshake	two door cinema club
5023	politik	coldplay
22081	hands of time	slightly stoopid
19922	how do you really feel	breakestra
26713	keep your eyes open	needtobreathe
26311	open your eyes	alter bridge
4345	hands	jewel
17180	so rare	george shearing
24791	accept yourself	the smiths
19719	open your eyes, you can fly	lizz wright



Conclusions

Unbalanced distribution of topics

Even though the romantic songs only represent 5% of the whole dataset I wanted to focus my analysis on these songs

Romance or violence?

The analysis hinted some obscure topics in romantic songs, but a further deeper analysis would be required

Less romance

We have seen how romantic songs have decreased through time to give way to other topics



Next steps

Build a classification model

To classify songs per genre and topic

Build a linear regression model

To see the (downward?) evolution of romantic songs. Is this the end of romance??

Further explore romantic songs

Explore other topics (like violence) and see if there is any correlation or if interesting information comes up



If I can't dance, I don't want to be in your revolution

Emma Goldman (1869-1940)

Thank you!